Auteursrecht en generatieve AI: een overzicht van 2024-2025 in vogelvlucht

Ontwikkelingen in rechtszaken van de VS en Europa

mr. F.F. Blokhuis[1]

Het auteursrecht wordt door AI-providers als een hinderlijk obstakel ervaren. Alle grote AI-providers hebben volgens onderzoek van de Danish Rights Alliance voor training van hun LLM modellen gebruik gemaakt van grote datasets, zoals LibGen, Books3 en Common Crawl. Deze datasets vormen een inbreuk op het auteursrecht. Er zijn tot op heden betrekkelijk weinig licenties gesloten. Bewust is gekozen om zoveel mogelijk data zonder toestemming te gebruiken voor de training van de AI-modellen. Vele rechthebbenden van miljoenen werken, zoals boeken, blogs, video's, muziek, beeld en software, eisen nu een verbod en/of schadevergoeding.[2]
Kunnen de AI-providers voor training van de LLM's deze werken ontleden en gebruiken, zonder enige compensatie of toestemming van de rechthebbenden?
De hamvraag is of de fair use-exceptie (in de Verenigde Staten (VS)) of de uitzonderingen voor Text and Data Mining (TDM), hiervoor een oplossing bieden.
Tegelijkertijd wordt er in Engeland een verhit debat gevoerd over het wetsvoorstel waarmee training op basis van auteursrechtelijke werken is toegestaan, tenzij er sprake is van een rechtsgeldig opt-out. De House of Lords heeft al vier keer een wetsvoorstel van de regering weggestemd. De Lords beschermen daarmee de artiesten tegen AI-training. Een vergelijkbaar debat is in Europa niet gevoerd, omdat in 2019 - toen de TDM-excepties in de DSM-richtlijn werden opgenomen - weinigen de impact van generatieve AI-systemen konden voorzien.[3]
In dit artikel worden een aantal uitspraken omtrent fair use en TDM besproken.

1. Inleiding

De voiceclone van Herman van Veen liet de speakers trillen: 'Wie heeft mijn stem gestolen?'

Dergelijke muziek wordt gemaakt met machines als Suno en Udio, die getraind zijn op – naar schatting[4] – honderdduizenden uren muziek. Deze machines zijn generatieve AI-modellen, die als output muziek heeft, waarvan niet (goed) te horen is dat die door generatieve AI zijn gemaakt. Er is zelfs al een virtuele band op Spotify, The Velvet Sundown met meer dan 1 miljoen! luisteraars per maand. Die muziek concurreert direct met echte musici.

Mag dat?

Of het een auteursrechtinbreuk is, is een vraag waar wereldwijd meer dan 75 rechtbanken en hoven zich over moeten buigen. Alleen al in de VS lopen er nu al meer dan 59 zaken tegen providers van AI-modellen of AI-systemen enkel over auteursrecht.[5]

2. Auteursrecht en AI-training

2.1. Trainen is verveelvoudigen

AI-providers gebruiken de datasets voor de training van hun AI-modellen. AI-providers willen veel data om hun modellen te trainen en licentieovereenkomsten zijn duur. Daarom gebruiken diverse AI-providers ook veel op internet publiek beschikbare data zonder toestemming van de rechthebbenden. Daarbij is het samenstellen van datasets (het scrapen) niet altijd een handeling die door de AI-provider is uitgevoerd. Er zijn reeds diverse datasets beschikbaar voor AI-training, los van de vraag of die datasets ongeoorloofde verveelvoudigingen (in de breedste zin van het woord) bevatten, waarvan de bekendste Common Crawl, Books3 en LibGen zijn.[6]

Om vast te stellen of er in de trainingsfase inbreuk is gepleegd, is het van belang om vast te stellen of er auteursrechtelijke verveelvoudigingen zijn gemaakt.

De trainingsfase kan weer verdeeld worden in pretraining, training en finetuning. De maatregelen die een AI-provider neemt om ervoor te zorgen dat de output geen inbreuk maakt, of geen onrechtmatige content genereert noemt men alignment. De align-

1. Fulco Blokhuis is advocaat bij Boekx Advocaten te Amsterdam.
2. rettighedsalliancen.dk/wp-content/uploads/2025/03/Report-on-pirated-content-used-in-training-of-AI.pdf
3. www.bbc.com/news/articles/clyrgv2n190o
4. Beide bedrijven geven geen informatie over de datasets die ze voor training hebben gebruikt.
5. Chatgptiseatingtheworld.com
6. Veel van die datasets bevatten werken zonder toestemming van de rechthebbenden.

ment moet er voor zorgen dat output geen inbreuk maakt.

Ik ga er voor nu van uit dat er tijdens training auteursrechtelijke verveelvoudigingen worden gemaakt. In het eerste Infopaq-arrest oordeelde het HvJ EU dat een "data-capture-procedé" (in casu: het scannen en automatisch doorzoeken van persartikelen door een attenderingsdienst) als reproductie in de zin van art. 2 Auteursrechtrichtlijn aangemerkt kan worden.[7] En in de meeste uitspraken die nu zijn gewezen draait het vooral om de reikwijdte van de uitzonderingen, en niet primair om de vraag of er sprake is van reproducties.

Voor een meer diepgaande analyse van het trainingsproces verwijs ik naar het uitstekende artikel 'The heart of the matter: copyright, AI-training, and LLM's' van Daniel Gervais.[8]

2.2. Licenties van datasets

Nu er bij AI-training miljoenen auteursrechtelijke verveelvoudigingen worden gemaakt, is er sprake van een auteursrechtinbreuk als dat zonder toestemming gebeurt.

De markt voor licenties voor AI-training is nog beperkt, maar komt langzamerhand op stoom.

Een aantal grote uitgeverijen heeft licentieovereenkomsten gesloten met AI-providers om hun data te gebruiken voor het trainen van AI-modellen. OpenAI heeft overeenkomsten gesloten met verschillende nieuwsorganisaties en uitgevers, waaronder The Associated Press, de uitgever van The Wall Street Journal, Le Monde, Financial Times en Axel Springer.[9]

Andere positieve voorbeelden van AI-modellen die met toestemming van rechthebbenden zijn getraind: GTP-NL voor Nederlandse tekst, Moonlit.ai voor video of het Zwitserse Apertus, het volledige open source, meertalige LLM.[10]

3. Wettelijke uitzonderingen en beperkingen

Toestemming voor training is uiteraard niet nodig als er een uitzondering van toepassing is. In de VS geldt de flexibele open uitzondering van fair use die is vastgelegd in Artikel 107 van de Copyright Act 1976.[11]

3.1. Fair use in de VS

De website chatgptiseatingtheworld.com volgt de inmiddels bijna 60 rechtszaken in de VS tegen AI-providers op grond van auteursrecht op de voet.

In alle rechtszaken over AI-training wordt fair use als verweer aangevoerd. Daarbij wordt er naar vier criteria gekeken.

A. Doel en karakter van het gebruik
B. De aard van het auteursrechtelijk beschermde werk
C. Hoeveel van het werk is er gebruikt?
D. Het effect dat het vermeende inbreukmakende gebruik heeft gehad op het vermogen van de auteursrechthebbende om zijn originele werk te exploiteren, ook wel *market dillution* of *market effect* genoemd.

3.2. Jurisprudentie:

Thomson Reuters Westlaw – ROSS

Van de vijftig rechtszaken in de VS die zijn aangespannen tegen AI-Providers zijn er tot nu drie waarin een interessant tussenvonnis is gewezen over fair use.

De eerste uitspraak over fair use in dit verband is de zaak tussen uitgeverij Thompson Reuters en legal tech onderneming Ross. De zaak die al in 2020 startte ging over het gebruik door ROSS Intelligence van 2.243 auteursrechtelijk beschermde samenvattingen van uitspraken om haar AI-model te trainen, zonder een licentie van rechthebbende Thomson Reuters.

Thomson Reuters is de exploitant van Westlaw, een toonaangevend juridisch researchplatform. Westlaw bevat vrijwel alle jurisprudentie uit de VS en bevat van elke uitspraak van zelfgeschreven headnotes – korte samenvattingen van de juridische kernpunten in die zaak. Daarnaast ordent Westlaw die headnotes in een genummerd Key Number System om gerelateerde uitspraken over hetzelfde juridische onderwerp te groeperen. Deze door jurist-redacteuren gemaakte toevoegingen bieden juristen snelle inzichten in rechtspraak.

ROSS Intelligence was een legal tech startup (opgericht rond 2014) die een AI-gedreven juridisch zoekinstrument wilde bouwen. Het idee was dat advocaten in natuurlijke taal vragen konden stellen en relevante jurisprudentie als antwoord kregen. ROSS vroeg Thomson Reuters tevergeefs om een licentie.

7. HvJ EU 16 juli 2009, C-5/08 (Infopaq I), r.o. 64
8. Gervais, Daniel J. and Shemtov, Noam and Marmanis, Haralambos and Zaller Rowland, Catherine, The Heart of the Matter: Copyright, AI Training, and LLMs (September 21, 2024). Beschikbaar via SSRN: https://ssrn.com/abstract=4963711 or http://dx.doi.org/10.2139/ssrn.4963711
9. www.businessinsider.nl/nu-ai-makers-op-zoek-zijn-naar-data-om-ai-op-te-trainen-lopen-ze-juist-tegen-meer-rechtszaken-aan/
10. ethz.ch/en/news-and-events/eth-news/news/2025/09/press-release-apertus-a-fully-open-transparent-multilingual-language-model.html
11. Artikel 107 van de Copyright Act 1976.

Vervolgens heeft ROSS het bedrijf LegalEase Solutions ingehuurd in om zogeheten "Bulk Memos" te maken. Dit waren grote sets juridische vragen met bijbehorende antwoorden, over uiteenlopende rechtsproblemen. Advocaten bij LegalEase gebruikten Westlaw's headnotes als referentie om die vragen en antwoorden op te stellen (met instructie ze niet letterlijk te kopiëren). De Bulk Memos dienden als trainingsmateriaal voor ROSS' AI-algoritme. In feite had ROSS zo toegang tot de essentie van Westlaw's gecureerde inhoud, zonder formeel iets van Thomson Reuters te kopen.

In die zaak heeft een federale rechter in Delaware (in *summary judgment*) bepaald dat de meeste samenvattingen (de zgn headnotes) van de (niet beschermde) jurisprudentie auteursrechtelijk beschermd waren.[12] De rechtbank oordeelde dat de headnotes "have more than the minimal spark of originality required for copyright validity. But the material is not that creative." Ook oordeelde hij dat het fair use verweer niet opging. Dat was opvallend, want in 2023 had dezelfde rechter dit verweer erkend en naar trial verwezen, zodat een jury er een oordeel over moest geven. Daar kwam hij nu dus van terug. De rechtbank redeneerde dat de AI-tool van Ross direct concurreert met Westlaw, wat neerkomt op commercieel gebruik (factor 4).

Het gebruik van de Westlaw Headnotes voor training van het systeem van ROSS was dus een auteursrechtinbreuk.[13]

Ross's gebruik van de Headnotes was ook niet transformatief omdat Ross's AI-tool niet een ander doel of karakter heeft dan Westlaw van Thomson Reuters.[14] Belangrijk hierbij is dat de rechtbank benadrukte dat Ross's AI geen GenAI is die zelf nieuwe content of samenvattingen genereert.
Ross heeft inmiddels hoger beroep ingesteld tegen deze beslissing.

Kadrey – Meta en Bartz – Anthropic: twee keer een Pyrrus overwinning voor fair use

Eind juni 2025 werden er twee uitspraken gedaan in de Kadrey-Meta en Bartz v Anthropic zaken, die veel aandacht kregen. Beide zaken gingen vooral over de input. De feiten in beide zaken komen aardig overeen. In beide zaken hadden de AI-providers Meta en Anthropic zonder toestemming van auteurs miljoenen boeken gebruikt voor AI-training.

Bartz - Anthropic
Anthropic had de dataset Books3 heeft gedownload, een online bibliotheek met bijna 200.000 boeken die was samengesteld uit illegaal gekopieerde exemplaren. Anthropic heeft daarna ten minste zeven miljoen extra exemplaren van boeken gedownload uit repositories op basis van illegale bronnen (waarvan Anthropic wist dat ze illegaal waren), waaronder boeken van de eisende auteurs. Opeenvolgende versies van haar LLM Claude zijn getraind op basis van deze boeken.

In 2024 deed Anthropic het anders. Zij kocht voor miljoenen papieren boeken, sneed deze op maat en scande de pagina's. Daarna werden de papieren versies weggegooid. Zij legde een onderzoekbibliotheek aan met al deze kopieën. Diverse van deze boeken werden opnieuw gekopieerd om trainingskopieën te maken. De trainingskopieën werden achtereenvolgens gekopieerd om te worden bewerkt voor training, "getokeniseerd" en "gecomprimeerd" tot een bepaalde getrainde LLM.

In de Kadrey-Meta zaak hebben dertien boekauteurs Meta gedagvaard voor het gebruik van hun boeken voor training van het Llama taalmodel. Daarbij heeft Meta bewust gebruik gemaakt van de Libgen dataset, een omvangrijke dataset met meer dan 7.5 miljoen boeken en 81 miljoen research papers, ook van Nederlandse auteurs.[15]

Een ander aspect van deze casus is dat Meta de Libgen dataset ook weer via p2p openbaar heeft gemaakt.

De rechters van het Northern District van California komen in beide zaken tot het oordeel dat er sprake is van fair use bij training van AI, maar de motivering verschilt aanzienlijk.

In Bartz -Anthropic[16] (C 24-05417 WHA) oordeelde rechter Alsup dat het doel en karakter van het gebruik van de boeken "bij uitstek transformatief" is. De rechter vergeleek AI-training met een lezer die schrijver wil worden en boeken leest, niet om ze te kopiëren en te vervangen, maar om iets anders te creëren.

Met betrekking tot vierde factor (*market dillution*) hadden de auteurs aangevoerd dat de LLM's voor een explosie van concurrerende werken zouden zorgen. De rechter was echter van oordeel dat die klacht niet anders was dan wanneer zij zouden klagen dat het trainen van schoolkinderen om goed te schrijven zou leiden tot een explosie van concurrerende werken.

Rechter Alsup concludeerde echter met betrekking tot het gebruik van de illegale kopieën dat het fair use verweer niet opging. In theorie zou dat betekenen dat Anthropic $ 150.000 per werk zou moeten betalen. Maal 7 miljoen werken komt dat neer op een schade die 17 keer hoger is dan de waarde van Anthropic op dat moment werd geschat.

12. www.ded.uscourts.gov/sites/ded/files/opinions/20-613_5.pdf
13. *Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc.*, No. 1: 20-CV-613-SB, at 14, 16 (D. Del. Feb. 11, 2025).
14. *Id.* at 17 (citing *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 529 (2023))
15. www.theatlantic.com/technology/archive/2025/03/search-libgen-data-set/682094/
16. storage.courtlistener.com/recap/gov.uscourts.cand.434709/gov.uscourts.cand.434709.231.0_4.pdf

Inmiddels is deze zaak geschikt voor 1.5 miljard dollar, wat neerkomt op 3.100 dollar per boek.

Kadrey- Meta

In Kadrey- Meta (Case No. 23-cv-03417-VC) kwam rechter Chhabria[17] tot een vergelijkbaar oordeel met betrekking tot de eerste drie factoren. Maar in de inleiding, die het lezen meer dan waard is, reageert hij op de uitspraak van Alsup van twee dagen daarvoor. Hij stelt dat:

"*using books to teach children to write is not remotely like using books to create a product that a single individual could employ to generate countless competing works with a miniscule fraction of the time and creativity it would otherwise take.*"[18]

Over de vraag of AI-training een inbreuk zal zijn is schrijft hij dat dat in veel zaken het geval zal zijn.

> *"What copyright law cares about, above all else, is preserving the incentive for human beings to create artistic and scientific works."*[19]

Hij concludeert dat de rechthebbenden gecompenseerd moeten worden als hun werken worden gebruikt voor training van AI-modellen.

> *"These products are expected to generate billions, even trillions, of dollars for the companies that are developing them. If using copyrighted works to train the models is as necessary as the companies say, they will figure out a way to compensate copyright holders for it."*[20]

In de materiele beoordeling van de vierde factor stelt hij dat market dilution heel relevant is.

> *"This case, unlike any of those cases, involves a technology that can generate literally millions of secondary works, with a miniscule fraction of the time and creativity used to create the original works it was trained on. No other use—whether it's the creation of a single secondary work or the creation of other digital tools—has anything near the potential to flood the market with competing works the way that LLM training does." "*[21]

De eisers hebben geen betekenisvol bewijs geleverd voor market dilution en dus slaagt het fair use verweer van Meta.

Het vonnis leest alsof Chhabria dit bijna jammer vindt:

> *"In cases involving uses like Meta's, it seems like the plaintiffs will often win, at least where those cases have better-developed records on the market effects of the defendant's use. No matter how transformative LLM training may be, it's hard to imagine that it can be fair use to use copyrighted books to develop a tool to make billions or trillions of dollars while enabling the creation of a potentially endless stream of competing works that could significantly harm the market for those books. And some cases might present even stronger arguments against fair use."*[22]

Het vergt niet veel moeite om te bedenken welke creatieve beroepsgroepen harder geraakt worden door de AI-slop dan de Kadrey auteurs: Stemacteurs, vertalers, stockfotografen, musici, copywriters, reclamemakers en softwareontwikkelaars.

Kortom, drie uitspraken met drie verschillende uitkomsten. Maar alle drie de rechters maken duidelijk dat het fair use-verweer vaak zal sneuvelen.

4. De uitzonderingen in de EU: de TDM-excepties

In Europa kennen we een gesloten systeem van excepties. Lidstaten mogen kiezen welke excepties van artikel 5 van de Auteursrechtrichtlijn (2001/29/EG) ze implementeren. Toepassing van de excepties moet worden getoetst aan de driestappentoets (artikel 5 lid 5 van de Auteursrechtrichtlijn). Deze toets bepaalt de grenzen waarbinnen uitzonderingen en beperkingen op auteursrechten mogen worden toegepast. De toets bestaat uit drie stappen:

i. De exceptie moet beperkt blijven tot specifieke situaties met een duidelijk maatschappelijk belang.
ii. De exceptie mag de normale commerciële exploitatie van het werk niet belemmeren.
iii. De rechten van de auteursrechthebbende mogen niet onredelijk worden geschaad.

Uit de jurisprudentie van het Hof volgt ook dat afweging met andere grondrechten binnen de excepties moet gebeuren.[23] Het Hof liet in de drie arresten van 29 juli 2019 geen zelfstandige uitzonderingen op auteursrechten toe buiten de uitputtende lijst van artikel 5 Auteursrechtrichtlijn. Het evenwicht tussen de verschillende grondrechten moet worden gevonden binnen de bestaande uitzonderingen. Een grondrecht an sich kan niet ingeroepen worden ter rechtvaardiging van een inbreuk op het auteursrecht.

In de DSM-richtlijn zijn twee artikelen geïntroduceerd die voor AI-training zeer relevant zijn. Het gaat om artikel 3 en 4 die derden het recht geeft om

17. storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.598.0.pdf
18. Case 3:23-cv-03417-VC Document 598 Filed 06/25/25 Pagina 3;
19. Pagina 1;
20. Pagina 4;
21. Pagina 32;
22. Pagina 39;
23. HvJ 29 juli 2019, nr. C-469/17, ECLI:EU:C:2019:623, Funke Medien/Bundesrepublik Deutschland; HvJ 29 juli 2019, nr. C-476/17, ECLI:EU:C:2019:624, Pelham e.a./Hütter e.a.; HvJ 29 juli 2019, nr. C-516/17, ECLI:EU:C:2019:625, Spiegel Online/Beck.

werken zonder toestemming van de rechthebbenden te gebruiken voor tekst en data mining.[24]

Ten eerste artikel 3 Digital Single Market richtlijn (2019/790/EG). Dit artikel geeft onderzoeksinstellingen het recht om reproducties te maken voor tekst en datamining doeleinden. Het geeft deze partijen echter niet het recht om ook de reproducties openbaar te maken.

Artikel 4 ziet op commercieel gebruik. Dit artikel is ooit door de Nederlandse delegatie voorgesteld.[25] Rechthebbenden hebben daarvoor de mogelijkheid om van een opt out recht gebruik te maken, waarmee ze hun werken kunnen uitsluiten van TDM (en dus voor AI-training). Beide uitzonderingen zijn met betrekkelijk weinig debat aangenomen in de DSM-richtlijn. In de AI Act[26] is echter bepaald dat deze artikelen ook van toepassing zijn op AI-training.

De reikwijdte van diverse begrippen uit deze artikelen staan ter discussie. Twee Duitse onderzoekers hebben betoogd dat deze uitzonderingen helemaal niet geschikt zijn voor training van AI.[27] Aan de andere kant van het spectrum heeft professor Margoni beargumenteerd dat rechtmatige toegang breed moet worden opgevat. Hij vindt: als het maar op internet staat, ook zonder toestemming van de rechthebbende, dan mag het voor wetenschappelijke doeleinden worden gebruikt voor training.[28] Het is wachten op de eerste prejudiciële vragen hierover: vragen als wat is rechtmatig toegang? Wat valt er onder een onderzoeksinstelling? Waar moet een rechtsgeldige opt-out verklaring aan voldoen?

De European Copyright Society heeft in haar Opinie van februari 2025[29] de Europese Unie er op gewezen dat de onzekerheid en verschillende open vragen met betrekking tot TDM snel behandeld moeten worden. Zo is de ECS terecht van mening dat de reikwijdte van de TDM excepties onderzocht en geanalyseerd moet worden. Dit omvat dat er een antwoord moet komen over de vraag of er handelingen van reproductie of een mededeling aan het publiek plaatsvindt en welke actoren aansprakelijk zijn voor dergelijke handelingen. Daarbij moet de mogelijkheid van commercieel gebruik van modellen die zijn getraind voor wetenschappelijk onderzoek worden onderzocht. Ook is belangrijk hoe de uitoefening van de opt-outmogelijkheid van artikel 4 kan worden uitgeoefend, alsmede het aspect van rechtmatig toegankelijke bronnen voor de uitzondering voor onderzoek waarin artikel 3 van de DSM-richtlijn voorziet.

5. Driestappentoets en het formaliteitenverbod van de Berner Conventie

Verschillende personen,[30] waaronder Axel Voss, rapporteur van het Europees Parlement en Prof Lucchi[31], die een uitvoerig rapport schreef op verzoek van de juridische commissie van het Europees Parlement, hebben betoogd dat het opt-outregime van art 4 TDM in strijd is met diverse bepalingen van de Berner Conventie en de driestappentoets van art 5 lid 5 van de InfoSoc Richtlijn.

Axel Voss schreef in zijn ontwerp rapport van juni 2025[32] dat de verwijzing in de AI Act naar de TDM excepties geen goede en proportionele oplossing bood.

Krachtens artikel 5 lid 2 van de Conventie geldt een absoluut formaliteitenverbod: het genot en de uitoefening van de door de Conventie beschermde rechten mag niet aan een formaliteitseis worden onderworpen. Het opt-out model dwingt makers echter om actie te ondernemen om hun werk te beschermen tegen AI-traning.

Sommigen betogen dat deze toets ook de Europese TDM-excepties kunnen torpederen. Maar dat is na het Kwantum/Vitra-arrest[33], waarin door het Europese Hof werd bepaald dat het reciprociteitsbeginsel door het Unierecht opzij kan worden gezet, nog maar de vraag.
Bovendien wordt het Amerikaanse systeem, waarbij makers hun werken bij het Copyright Office moeten deponeren, om het te kunnen handhaven, ook niet in strijd met art. 5 BC geacht.

5.1. Oplossing

Senftleben, Voss en Lucchi pleiten allemaal voor een vergoeding die door collectieve beheersorganisaties geïnd en verdeeld zou moeten worden. Die gedachte is in lijn met andere aanpassingen van het auteurs-

24. Zie verder uitgebreid P.B. Hugenholtz in AMI 2019/5, *Artikelen 3 en 4 DSM-richtlijn: tekst- en datamining*: tekst- en datamining https://www.ivir.nl/publicaties/download/AMI_2019_5.pdf
25. P.B. Hugenholtz, Artikelen 3 en 4 DSM-richtlijn: tekst- en datamining, AMI 2019/5
26. Verordening (EU) 2024/1689 van het Europees Parlement en de Raad van 13 juni 2024 tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie en tot wijziging van de Verordeningen (EG) nr. 300/2008, (EU) nr. 167/2013, (EU) nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 en (EU) 2019/2144, en de Richtlijnen 2014/90/EU, (EU) 2016/797 en (EU) 2020/1828 (verordening artificiële intelligentie) (Voor de EER relevante tekst)
27. *Urheberrecht und Training generativer KI-Modelle technologische und juristische Grundlagen,* Tim W. Dornis und Sebastian Stober August 2024, papers.ssrn.com/sol3/papers.cfm?abstract_id=4946214
28. T. Margoni, *TDM and Generative AI: Lawful Access and opt-outs,* Forthcoming in Auteurs & Media 2024; papers.ssrn.com/sol3/papers.cfm?abstract_id=5036164
29. https://europeancopyrightsociety.org/opinions/
30. Zie Senftleben, Martin, TDM, GenAI and the Copyright Three-Step Test (24 juni 2025). SSRN: https://ssrn.com/abstract=5373903
31. Study verzocht door de JURI Comissie van het Europese Parlement, Prof Lucchi: Generative AI and Copyright zie www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025
32. www.europarl.europa.eu/doceo/document/JURI-PR-775433_EN.html
33. HvJ EU 24 oktober 2024, C-227/23 - Kwantum Nederland and Kwantum België;

recht aan technologische ontwikkelingen, zoals bijvoorbeeld de thuiskopie, leenrecht, etc.

6. Jurisprudentie in Europa

6.1. Lopende procedures

Op 21 januari 2025 heeft de Duitse collectieve rechtenbeheerorganisatie GEMA een rechtszaak aangespannen tegen Suno Inc. bij de regionale rechtbank van München (Landgericht München). Suno en Udio waren in juni 2024 ook al aangeklaagd door een aantal majors (platenmaatschappijen). In die procedure hebben de AI-providers erkend dat beschermde muziek in de trainingsdata zat.

GEMA was ook al een procedure gestart tegen OpenAI voor het gebruik van songteksten. De uitspraak wordt op 11 november 2025 verwacht, maar uit berichtgeving komt naar voren dat de Rechtbank München gaat oordelen dat het beroep op de TDM exceptie niet opgaat.[34]

Inmiddels is er in Frankrijk een procedure gestart door Nationale Uitgeversbond (SNE), de nationale vakbond van auteurs en componisten (SNAC) en de Vereniging van Letterkundigen (SGDL) tegen Meta. In Engeland is er vlak voor publicatie van dit artikel een zeer uitvoerig vonnis gewezen in de procedure van Getty Images tegen Stability AI. In Denemarken zijn persuitgevers een procedure tegen OpenAI gestart.[35] OpenAI is nu de provider die het meest gedagvaard is.[36] In Europa zijn er prejudiciële vragen gesteld door een Hongaarse rechter in een procedure tussen uitgever Like en AI-provider Google.[37]

6.2. Like Company – Google Ireland C-250/25

De Hongaarse persuitgever Like Company exploiteert de website balatonkornyeke.hu met journalistieke artikelen met daaromheen reclame. In een van de auteursrechtelijk beschermde online publicaties van Like was een artikel verschenen dat berichtte dat Kozsó, een bekende Hongaarse zanger, nog steeds zijn droom koestert om in de buurt van het Balaton, het grootste meer van Hongarije, dolfijnen te houden in een aquarium.

Via Google's AI-model Gemini kon zonder toestemming samenvattingen van de perspublicaties worden gegenereerd. Op de prompt: "Kunt u een samenvatting maken in het Hongaars van het online artikel over de plannen van Kozsó voor het plaatsen van dolfijnen in het meer, dat is verschenen op balatonkornyeke.hu?" gaf Gemini (vroeger Bard) een uitvoerig antwoord, inclusief een samenvatting van de perspublicaties van Like. Het zou om dit artikel gaan:

balatonkornyeke.hu/kozso-nem-adja-fel-tovabbra-is-delfineket-szeretne-a-balatonhoz-telepiteni-a-nepszeru-enekes/

Like stapte naar de Hongaarse rechter. Volgens de uitgever bevatte de samenvatting van Gemini auteursrechtelijk beschermde inhoud, en dus een ongeoorloofde reproductie van het oorspronkelijke artikel en mededeling aan het publiek.

Like beriep zich op artikel 15 van de Europese richtlijn 2019/790, dat uitgevers van perspublicaties naburige rechten verleent. Zij stelde dat Google de inhoud van haar publicaties tussen herhaaldelijk zonder toestemming heeft gereproduceerd en beschikbaar heeft gesteld aan het publiek.

Google verweerde zich met de argumenten dat zij geen volledige kopieën van de teksten opsloeg, maar de informatie dynamisch verwerkt. Er zou informatie zijn "gehallucineerd". Het interessante verweer is dat er geen nieuw publiek is bereikt, omdat de AI-teksten net als de originele publicaties toegankelijk zijn voor alle internetgebruikers. En uiteraard werden de uitzonderingen van stal gehaald, zoals de tijdelijke reproductie en TDM.

6.3. De prejudiciële vragen over AI-training en AI-output

De Hongaarse rechter wil vervolgens weten wat de juridische grenzen zijn van het gebruik van de perspublicaties door een LLM. De rechter stelt vervolgens vragen aan het Europese Hof over persuitgeversrecht, mededeling aan het publiek en reproductie in het licht van training (zaak C-250/25):

De eerste vraag ziet op het begrip mededeling aan het publiek:

1) Moeten persuitgeversrecht en het recht van mededeling aan het publiek, zo worden uitgelegd dat de weergave van tekst die gedeeltelijk overeenstemt met de inhoud van webpagina's in de antwoorden van een LLM chatbot, moet worden beschouwd als mededeling aan het publiek?

Zo ja, is het dan relevant dat dit het resultaat is van een proces waarbij de chatbot uitsluitend op basis van de waargenomen patronen voorspelt wat het volgende woord gaat worden?

De volgende vragen zien op het reproductierecht t.o.v. training en output. Samengevat luiden ze zo:

34. aifray.com/breaking-openai-on-losing-track-in-german-copyright-case-brought-by-music-rights-collecting-society-over-song-lyrics-injunction-looms-large/
35. Getty Images vs Stability AI in Engeland, Gema vs OpenAI en Gema vs Suno in Duitsland en de belangrijkste uitgeverij- en auteursverenigingen tegen Meta in Frankrijk. In Nederland is vooral Stichting Brein aan het handhaven, maar dat heeft nog niet tot een procedure geleid.
36. Chatgptiseatingtheworld.com
37. Case C-250/25

2) Is het trainingsproces van een LLM een reproductie?
3) Als dat zo is, geldt dan de uitzondering van artikel 4 (commerciële TDM)?
4) Als een AI-chatbot een opdracht krijgt die overeenkomt met of verwijst naar tekst in een perspublicatie en de chatbot vervolgens een antwoord genereert waarin de inhoud van de perspublicatie geheel of gedeeltelijk wordt weergegeven, is dit dan een reproductie?[38]

6.4. Zijn de feiten wel goed genoeg voor scherpe antwoorden?

Tot zover lijkt het er op dat de belangrijkste auteursrechtelijke vragen eindelijk gesteld zijn aan de hoogste Europese rechter.

Wanneer echter nauwkeurig naar het originele artikel en de handelingen van Google wordt gekeken valt te betwijfelen of deze zaak nu de meest geschikte was.

Zo schrijft Andres Guadamuz dat het origine artikel slechts 579 woorden bevat, en uit samenvattingen van andere websites lijkt te bestaan.[39] Bij de artikelen is geen naam van een auteur vermeld. De foto van Koszo lijkt van zijn facebookpagina te komen. Het is zeer de vraag of het originele artikel een auteursrechtelijk werk is. Ook is de output van een prompt, de samenvatting die de LLLM maakt, in principe niet openbaar en dus geen mededeling aan een nieuw publiek. Het wordt enkel voor de prompteur gegeneerd. En het verschil meestal iedere keer.

De vraag is of de vragen over de reproductie een antwoord gaan geven op de belangrijkste auteursrechtvraag van deze tijd. Paul Keller wees er op dat het er eerder op lijkt dat het artikel geen onderdeel lijkt te zijn van de trainingsdata, maar via Retrieval Augmented Generation (RAG) door Google is opgevraagd van de website van Like Company.[40] Hij stelt dat het vrijwel zeker is dat het reeds getrainde model Bard (thans Gemini) de live inhoud van de website als input heeft gebruikt en deze vervolgens heeft verwerkt om de gevraagde samenvatting te produceren. Deze interpretatie wordt ook ondersteund door de uitleg van Google: *"Om gegevens te verzamelen, maakt [de chatbot] gebruik van de Google Search-database en kan hij in zijn antwoord een aangepaste versie van een artikel weergeven, indien de gebruiker de originele versie van het artikel al in zijn instructies heeft verstrekt."* Met andere woorden, na ontvangst van de prompt zocht de chatbot in de Google Search-index naar inhoud van de website waarnaar werd verwezen en produceerde vervolgens een samenvatting op basis van die inhoud. Er werd dus niet iets uit de trainingsdata samengevat.

Kortom, het zou beter zijn als er in een andere zaak, bijvoorbeeld GEMA-OpenAI prejudiciële vragen worden gesteld, zodat Het Europese Hof nog de ruimte krijgt om heldere duiding te geven op basis van scherpe feiten.

Desalniettemin zal er reikhalzend naar de beantwoording in Like – Google worden uitgekeken.

6.5. Uitspraken over de TDM-exceptie

Ondertussen zijn er al twee uitspraken in het kader van de TDM-exceptie.

Kneschke -LAION

De eerste uitspraak ging over een Duitse stock fotograaf Robert Kneschke die er via de website haveib eentrained.com/ achter kwam dat een foto van een aantal bejaarden die aan fitness deden gebruikt was in de dataset van non-profit organisatie LAION. De foto stond op bigstockphoto.com en was openbaar toegankelijk. Op de website stond een helder voorbehoud:

"RESTRICTIONS YOU MAY NOT:

Use automated programs, applets, bots or the like to access the bigstockphoto.com website or any content thereon for any purpose, including, by way of example only, downloading Content, indexing, scraping or caching any content on the website."

De foto kwam toch voor in de LAION dataset. LAION had een dataset gecreëerd met hyperlinks naar beeldbestanden en bijbehorende beschrijvingen, voortbouwend op een oudere dataset van Common Crawl. Voor het maken van deze dataset heeft LAION tijdelijk beeldbestanden gedownload om de beschrijvingen te verifiëren, waaronder een foto van Kneschke. LAION ontwikkelt zelf geen AI-modellen maar stelt de dataset gratis beschikbaar aan onderzoekers. Deze dataset was gebruikt door Stability AI voor training van haar AI-model. LAION werd deels gefinancierd door Stability AI.

Kneschke dagvaardde LAION voor het Landgericht in Hamburg. LAION deed een succesvol beroep op artikel 3 DSM. De rechtbank honoreerde dit punt en wees de vordering van Kneschke af.

LAION kan zich beroepen op artikel 60a van de Duitse auteurswet (implementatie van artikel 3 DSM-richtlijn). De rechtbank hanteert een ruime opvatting van 'wetenschappelijk onderzoek': het volstaat dat de activiteit uiteindelijk gericht is op kennisvermeerdering.

Echter, in een obiter dictum geeft de rechtbank en-

38. curia.europa.eu/juris/showpdf.jsf?text=\&docid=300681\&pageindex=0\&doclang=nl\&mode=req\&dir=\&occ=first\&part=1\&cid=5661670
39. www.technollama.co.uk/first-case-on-ai-and-copyright-referred-to-the-cjeu;
40. legalblogs.wolterskluwer.com/copyright-blog/do-ai-models-dream-of-dolphins-in-lake-balaton/

kele overwegingen over artikel 4 DSM-richtlijn, dat TDM in algemene zin toestaat tenzij rechthebbenden uitdrukkelijk een voorbehoud maken. Voor de rechtsontwikkeling is verheldering hiervan zeer wenselijk.

De rechtbank verwijst naar de AI Act als bevestiging dat artikel 4 DSM-richtlijn van toepassing is op het gebruik van werken voor trainingsdoeleinden. De rechtbank past vervolgens de driestappentoets toe door te stellen dat de normale exploitatie van werken niet wordt geschaad door de totstandkoming van de dataset.

De rechtbank oordeelt voorts dat het voorbehoud door bigstockphoto als machine leesbaar kan worden beschouwd. Daar zijn door Hugenholtz een paar interessante argumenten tegenin gebracht.[41]

Hugenholtz bekritiseert de opvatting van de rechtbank dat een licentienemer (het fotoagentschap) als 'rechthebbende' kan kwalificeren voor het maken van een voorbehoud onder artikel 4 DSM-richtlijn. Hij meent dat alleen auteursrechthebbenden of door hen gemachtigden dit zouden moeten kunnen doen. Hugenholtz is het oneens met de interpretatie dat een in natuurlijke taal gesteld voorbehoud als 'machine-leesbaar' kan worden beschouwd. Hij stelt dat de crawlers die websites doorzoeken eenvoudige programma's zijn zonder AI-capaciteiten.

De interpretatie van de rechtbank over machine-leesbaarheid staat volgens Hugenholtz ver af van de zich ontwikkelende praktijk, zoals blijkt uit de recente AI Code of Practice, waarin het Robots Exclusion Protocol (robots.txt) als prototype voor machine-leesbare opt-outs wordt gezien. Overigens zijn er meerdere geleerden die het gebruik van robots.txt niet geschikt vinden.

Kneschke heeft hoger beroep ingesteld. Een nieuwe kans voor (aanvullende) prejudiciële vragen dus. De uitspraak wordt in december 2025 verwacht.

Howards Home

De tweede uitspraak betrof een uitspraak van rechtbank Amsterdam in een kwestie over persuitgeversrecht, auteursrecht en databankenrecht. [42] De zaak betrof geen AI, maar is relevant in verband met het TDM-verweer. Howards Home had RSS feeds van diverse kranten (DPG Media, Mediahuis en NRC (de Uitgevers) gebruikt voor een eigen nieuwsdienst. De signaleringen bevatten een hyperlink naar het relevante bericht, inclusief de titel, een korte beschrijving van maximaal 150 tekens en wanneer beschikbaar een thumbnail.

De vorderingen van de uitgevers stranden. Tegen de auteursrechtelijke vorderingen voerde Howards-Home aan dat zij een beroep kon doen op het citaatrecht en de commerciële TDM exceptie van art 4 DSM en 15o Aw.

Daarover overweegt de rechtbank dat Howards-Home alleen openbare teksten op geautomatiseerde wijze heeft geanalyseerd. Of de activiteiten van HowardsHome uberhaupt wel een TDM activiteit is wordt niet overwogen. Wel overweegt de rechtbank in 4.33 dat het voorbehoud in robot.text niet toereikend was. HowardsHome had aangevoerd dat daarmee alleen bepaalde AI-bots worden uitgesloten zoals GPTBot, ChatGPT-User, CCBOT en anthropic-ai.

Daarmee stond onvoldoende vast dat het auteursrecht op de websites van de Uitgevers in machinaal leesbare middelen op passende wijze is voorbehouden. Of de andere voorbehouden , die niet in het voor AI-training omstreden robot.txt stonden, [43] wel voldoende waren wordt niet meegewogen. In de Kneschke-LAION zaak werd een dergelijk voorbehoud wel meegewogen.

Dit toont aan hoe lastig het voor rechthebbenden kan zijn om een goed voorbehoud te maken. Veel auteursrechtelijke werken worden immers gescraped, voordat bekend was wie of waarvoor dat gebeurde. Daarnaast is onduidelijk waar de voorbehouden geplaatst moeten worden, of hoe dat voor gebruikers te achterhalen is. Ook blijft onduidelijk wat grenzen van rechtmatige toegang precies zijn.

7. Code of Practice

Dan zijn er nog ontwikkelingen op grond van de AI Act. Op 10 juli 2025 is de General-Purpose AI (GPAI) Code of Practice van de Europese Commissie gepubliceerd.[44] Een van de hoofstukken gaat over auteursrecht. Op grond van art 53 AI Act, moeten GPAI-providers, zoals OpenAI, een beleid opstellen.

In de Code of Practice staan een aantal maatregelen die de AI-provider moet nemen:

1. een auteursrecht policy opstellen en implementeren;
2. alleen rechtmatig toegankelijke content scrapen;
3. voldoen aan het opt out regime van art. 4(3);
4. voor output: maatregelen nemen om inbreukmakende output te voorkomen; en
5. een contact punt aanwijzen voor klachtenafhandeling.

Maatregel 2 ziet er op dat er geen paywalls omzeild mogen worden en dat Pirate websites die (PirateBay

41. Hugenholtz, Auteursrecht 2025/1 Nr. 2 - Landgericht Hamburg 27 september 2024, Kneschke/LAION
42. Rb. Amsterdam 30 oktober 2024, IEF 22332, IT 4672; ECLI:NL:RBAMS:2024:6563 (de Uitgevers tegen HowardsHome)
43. Zie uitvoerig over het TDM voorbehoud: A.P. Engelfriet & D.J.G. Visser, 'Werkt de mijnwerk opt-out voor mijn werk?', *Atr* 2024, p. 12 ev.
44. digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

bijvoorbeeld) herhaaldelijk inbreuk maken niet gescraped mogen worden. De EC zal een dynamische lijst daarvoor publiceren.

Maatregel 3 krijgt het meeste toelichting.

De webcrawlers moeten voldoen aan het "Robot Exclusion Protocol (robots.txt)," en daarop volgende versies die implementeerbaar zijn.

Daarnaast moet aan andere machine-readable protocols worden voldaan die rechten voorbehouden op grond van art 4 Rl 2029/790, en welke protocollen zijn aangenomen door zijn internationale of Europese standaardisering organisaties.

De AI-providers worden wel opgeroepen om hierover met rechthebbende in debat te gaan met als doel om passende standaarden en protocollen te ontwikkelen.

De vraag is of de opt-outs in EU-stijl ook geldt voor inhoud die buiten de EU wordt gehost.

Er is eerder ook al veel kritiek geweest op het Robots.txt format. Cloudflare is zelf met een systeem gestart wat crawling alleen toestaat als de websitehouder daarmee instemt (en mogelijk betaling vraagt). Dat systeem werkt niet met Robots.txt

8. Hoe verder?

Uit de besproken uitspraken lijkt te volgen dat het 'fair use'-verweer in de VS niet op zal gaan als rechthebbenden hun zaak goed voorbereiden. Het gebruik van datasets die door piraterij tot stand zijn gekomen is eenvoudigweg niet toegestaan. In Europa zijn de twee uitspraken over de uitleg van de grenzen van het TDM-verweer onvoldoende richtinggevend. Het is wachten op het Europese hof of Rechtbank München om meer helderheid te scheppen. De voortekenen uit München zijn dat ook het TDM-verweer zal worden afgewezen. Diverse schrijvers betogen dat een TDM-verweer niet op zou moeten gaan. Het zal mij niet verbazen als rechters (vaker) concluderen dat het trainen van AI-modellen een inbreuk is op het auteursrecht, en dat makers gecompenseerd moeten worden.

Ondertussen blijft het aantal rechtszaken tegen AI-providers maar toenemen. Het afstraffen van het gebruik van een illegale data set in Bartz Anthropic heeft tot meerdere nieuwe rechtszaken geleid. Het toont aan dat vele eisers overtuigd zijn dat deze technologische ontwikkeling in strijd is met het huidige auteursrecht.

Aan de andere kant zullen de AI-ontwikkelaars met hun miljarden investeringen dit tot de hoogste rechters procederen, of proberen te schikken. Dat eerste zal nog jaren kunnen duren en het risico is groot dat de gewenste duidelijkheid niet uit diverse uitspraken is af te leiden.

Universal Musi Group heeft recent aangekondigd met Udio samen te gaan werken en makers te ondersteunen.

De recente uitspraken zijn echter een belangrijk signaal dat het trainen op basis van auteursrechtelijke datasets niet zonder risico is. De verwachting is daarom dat er steeds meer licentiedeals zullen komen. Het blijft immers een apart fenomeen dat er miljarden in de AI-industrie worden gepompt, terwijl dat vooral aan rekenkracht en softwaredevelopers wordt besteed en niet aan de rechthebbenden wiens werken als grondstof voor de training van de AI-modellen is gebruikt.

Of en wanneer er een verplichte vergoeding voor makers middels een uitzondering op het auteursrecht wordt geïntroduceerd is afwachten. Het past in lijn met verschillende uitzonderingen die de afgelopen decennia in het auteursrecht zijn geïmplementeerd na technologische ontwikkelingen, zoals de thuiskopie, of de vergoedingen die geïnd worden door collectieve beheersorganisaties voor uitzending via radio of tv.

Het zijn niet alleen de AI-providers die risico lopen om inbreuk te maken. Ook gebruikers die een pdf of andere beschermde content in een AI-model uploaden om het te laten vertalen of samenvatten voor gebruik op werk worden niet gedekt door een uitzondering.[45] Toch gebeurt dat op grote schaal. Dat roept ook de vraag of daar niet een aan het reprorecht vergelijkbare vergoeding in het leven voor moet worden geroepen.

De stormachtige ontwikkeling in het gebruik van generatieve AI-tools wijst er op dat dit risico niet echt onderkend wordt. Het zal nog wel even duren voordat er echt duidelijke antwoorden zijn. Tenzij de wetgevers ingrijpen. Dat dient in Europa dan op Europees niveau geregeld te worden.

45. Visser, D.J.G.; Smit, M.A. (2024) Bedrijfsjuridische Berichten, 20024, 17 'Van thuiskopie naar thuisbewerking: is het toegestaan om met AI-systemen teksten van anderen voor eigen gebruik te bewerken, te vertalen of samen te vatten?'